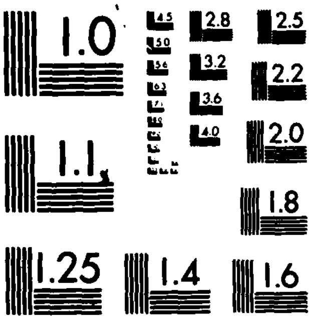
AD-A139 309 RESIDUALS IN NONLINEAR REGRESSION(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER R D COOK ET AL. JAN 84 MRC-TSR-2625 DAAG29-80-C-0041

1/1

UNCLASSIFIED F/G 12/1 NL

END
DATE
FILMED
4-84
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

1

MRC Technical Summary Report #2625

RESIDUALS IN NONLINEAR REGRESSION

R. D. Cook and C. L. Tsai

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

January 1984

(Received November 3, 1983)

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

-a-

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

RESIDUALS IN NONLINEAR REGRESSION

R. D. Cook and C. L. Tsai*

Technical Summary Report #2625
January 1984

ABSTRACT

The authors ~~We~~ employ a quadratic expansion to investigate the behavior of the ordinary residuals in nonlinear regression. In particular, they ~~we~~ derive quadratic approximations for the mean and variance of the ordinary residuals, and the covariances between the ordinary residuals and the fitted values. This investigation leads to the conclusion that the ordinary residuals can produce misleading results when used in diagnostic methods analogous to those for linear regression. Consequently, we suggest a new type of residual that overcomes many of the potential shortcomings of the ordinary residuals. is suggested

AMS (MOS) Subject Classification: 62J02

Key Words: Diagnostics, intrinsic curvature array, nonlinear regression, residuals

Work Unit Number 4 (Statistics and Probability)

*Department of Statistics and Operations Research, New York University, New York, NY 10006

Sponsored in part by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

Statistical methods for the analysis of experimental data are necessarily dependent on the specification of a model, a mathematical formula that describes the behavior of the data up to a few unknown parameters. Generally, a model can be visualized as

Datum (D) = Systematic component (S) + Random component (R)

or in abbreviated form $D = S + R$. The specification of a model often involves making assumptions, such as "the data are normally distributed," that may have little prior substantive support. Consequently, it becomes necessary to use the data to assess the adequacy of the model. Such assessments are extremely important in statistical analyses since erroneous assumptions can lead to erroneous conclusions (the mistaken conclusion that a drug is not carcinogenic could have devastating results).

Models that are linear in the unknown parameters, $\theta_1,\ldots,\theta_p$, have systematic components that can be expressed as

$$S = X_1\theta_1 + X_2\theta_2 + \ldots + X_p\theta_p$$

where the X_i's are nonrandom experimental variables whose values are known. Many methods of assessing model adequacy are available for such linear models. However, relatively little is known about how to assess model adequacy when S is <u>nonlinear</u> in the parameters; for example $S = \exp\{X_1\theta_1 + \ldots + X_p\theta_p\}$. The purpose of this paper is to provide a foundation for the development of methods for assessing the adequacy of models that are nonlinear in the parameters.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the authors of this report.

RESIDUALS IN NONLINEAR REGRESSION

R. D. Cook and C. L. Tsai*

1. INTRODUCTION

Diagnostic methods are useful for assessing the adequacy of assumptions underlying the modeling process and for identifying unexpected characteristics of the data that may seriously influence conclusions or require special attention. It is widely held that the diagnostic phase is an important part of any regression analysis.

A variety of diagnostic methods are available to aid in analyses based on linear regression models (Cook and Weisberg 1982 provide a review). For the most part, the development of these methods is dependent on a thorough study and characterization of the exact small sample behavior of a few fundamental building blocks such as the ordinary residuals and related statistics. The interpretation of standard residual plots, for example, depends on the knowledge that the expectations of the residuals are zero under a correct model.

In more complicated settings such as nonlinear regression, the exact small sample behavior of the corresponding building blocks is generally intractable so that some degree of approximation is necessary. In addition, the nonlinear regression problem involves new concerns that do not have counterparts in linear regression and thus that may require the development of new diagnostic methods.

Diagnostics for nonlinear regression can be constructed by using first-order extensions of analogous methods for linear regression (see, for example, Cook and Weisberg 1982). Generally, these diagnostics are based on the assumption that the usual tangent plane approximation to the solution locus is adequate,so that the nonlinear model is essentially linear in a neighborhood of the estimated parameters. While such diagnostic methods are certainly useful as first approximations and will often provide important information, a deeper analysis may be required for an adequate understanding of nonlinear regression.

*Department of Statistics and Operations Research, New York University, New York, NY 10006

Sponsored in part by the United States Army under Contract No. DAAG29-80-C-0041.

In this paper, we investigate properties of the ordinary residuals and related quantities from nonlinear regression. This investigation is based on the quadratic approximation of the ordinary residuals developed in section 2. In section 3, we derive informative expressions for the expectation and variance of the vector of ordinary residuals, and discuss why these residuals may not be an adequate basis for diagnostics methods. In section 4, we propose a new type of residual for use in nonlinear regression. It is shown that these new residuals overcome many of the failings of the ordinary residuals and that they can be used in much the same way as the ordinary residuals from linear regression. In the remainder of this section, we establish notation and briefly review relevant background material.

The standard nonlinear regression model can be represented as

$$y_i = f(x_i,\theta) + \varepsilon_i, \quad i = 1,2,\ldots,n \tag{1}$$

where x_i represents a vector of known explanatory variables associated with the i-th observable response y_i, θ is a $p \times 1$ vector of unknown parameters, the response function f is assumed to be known, continuous and twice differentiable in θ, and the errors ε_i are assumed to be independent, identically distributed normal random variables with mean 0 and variance σ^2. For this model, the maximum likelihood estimator $\hat{\theta}$ of θ can be found by minimizing the objective function

$$J(\theta) = \sum_{i=1}^{n} (y_i - f(x_i,\theta))^2 \tag{2}$$

Kennedy and Gentle (1980) discuss computational methods for obtaining $\hat{\theta}$; for our purposes we assume that $\hat{\theta}$ is available. The asymptotic behavior of $\hat{\theta}$ is investigated by Wu (1981) who provides additional references. The usual estimator of σ^2 is $s^2 = J(\hat{\theta})/(n - p)$.

For notational convenience, let $f_i = f(x_i,\theta)$, $i = 1,2,\ldots,n$, and let V denote the $n \times p$ matrix with elements $f_i^r = \partial f_i/\partial\theta_r$, $i = 1,2,\ldots,n$, $r = 1,2,\ldots,p$. Unless indicated otherwise, all derivatives are evaluated at the true parameter values. Various quadratic expansions used in the following sections involve the $p \times p$ matrices W_i, i = 1,2,...,n; the elements of W_i are $f_i^{rs} = \partial^2 f_i/\partial\theta_r\partial\theta_s$, r,s, = 1,2,...,p. These matrices can he written conveniently in an n × p × p array W (Bates and Watts, 1980). The kj-th "column" of W is the kj-th second derivative vector with elements f_i^{kj}, i = 1,2,...,n, while the i-th face W_i of W is the p × p matrix consisting of the i-th elements of the second derivative vectors.

2. ORDINARY RESIDUALS

The $n \times 1$ vector of ordinary residuals e can be written as

$$e = Y - f(\hat{\theta}) \tag{3}$$

where Y and $f(\hat{\theta})$ are $n \times 1$ vectors with elements y_i and $f(x_i, \hat{\theta})$, $i = 1,2,\ldots,n$, respectively. The vector e is of course a function of the errors ε_i, $i = 1,2,\ldots,n$. To investigate the properties of e, we use the quadratic expansion of the right side of (3) obtained by ignoring all terms that involve cubic and higher powers in the errors. This method of approximation is closely related to that in Cox and Snell (1968), Box (1971) and Clarke (1980).

The standard quadratic expansion of $f(\hat{\theta})$ about the true value θ^* is

$$f(\hat{\theta}) \approx f(\theta^*) + V(\hat{\theta} - \theta^*) + \frac{1}{2}(\hat{\theta} - \theta^*)^T W(\hat{\theta} - \theta^*) \tag{4}$$

where W is the $n \times p \times p$ array with i-th face W_i, $i = 1,2,\ldots,n$. Multiplication involving three dimensional arrays is defined as in Bates and Watts (1980) so that the third term of (4) is an $n \times 1$ vector with elements $(\hat{\theta} - \theta^*)^T W_i (\hat{\theta} - \theta^*)/2$, $i = 1,2,\ldots,n$. Substituting (4) into (3) we obtain the initial representation

$$e \approx \varepsilon - V\phi - \frac{1}{2}\phi^T W \phi \tag{5}$$

where for notational convenience $\phi = \hat{\theta} - \theta^*$. Since cubic and higher powers in the ε_i's are to be ignored, the standard first-order approximation $\phi \approx (V^TV)^{-1}V^T\varepsilon$ (Cox and Snell, 1968) can be substituted into the third term of (5):

$$\phi^T W \phi \approx \varepsilon^T V(V^TV)^{-1} W (V^TV)^{-1} V^T \varepsilon \tag{6}$$

To evaluate the second term of (5), we require a quadratic approximation of ϕ. Such an approximation can be obtained from the quadratic expansion of the likelihood equations about the true value θ^* (Cox and Snell, 1968). As shown in the Appendix, this yields

$$V\phi \approx P_1\varepsilon + V(V^TV)^{-1} \sum_{i}^{n} (\varepsilon^T \ell_i) W_i (V^TV)^{-1} V^T \varepsilon - \frac{1}{2} P_1\{\varepsilon^T V(V^TV)^{-1} W (V^TV)^{-1} V^T \varepsilon\} \tag{7}$$

where $P_1 = V(V^TV)^{-1}V^T$ is the projection operator for the column space of V and ℓ_i is the i-th column of $I - P_1$.

In the remainder of this paper, we use $C(F)$ to indicate the column space of the matrix F. Thus for example, the tangent plane at θ^* is the affine subspace $f(\theta^*) + C(V)$. The orthogonal complement of $C(F)$ will be denoted by $C'(F)$.

Expressions (6) and (7), which form the essential ingredients of (5), can be expressed more informatively in terms of the QR-decomposition $V = QR$ of V. Here Q is an $n \times n$ matrix with orthogonal columns and $R^T = (L^T, 0)$ where L is a $p \times p$, nonsingular, upper triangular matrix. Partition $Q = (U,N)$ where U is $n \times p$. The columns of U form an orthonormal basis for $C(V)$ and the columns of N form an orthonormal basis for $C'(V)$ so that $C(N) = C'(V)$. In terms of the transformed coordinates $\tilde{\theta} = L(\theta - \theta^*)$, the first and second derivative vectors are given by the columns of U and $\tilde{W} = L^{-T}WL^{-1}$. The i-th face of $\tilde{W}$ is simply

$$\tilde{W}_i = L^{-T}W_iL^{-1} \qquad (8)$$

Using the QR-decomposition to simplify (6) and (7), and substituting the resulting expressions into (5) gives

$$e = N\eta - U\sum_i^n (\varepsilon^T \ell_i)\tilde{W}_i\tau - \frac{1}{2} NN^T(\tau^T\tilde{W}\tau) \qquad (9)$$

where $(\tau^T, \eta^T) = Q^T\varepsilon$ is the vector of rotated errors. The components of $Q^T\varepsilon$ are, of course, independent and follow the same distribution as that assumed for ε.

Some additional discussion of (9) should prove useful. First, the ab-th element of the $p \times p$ matrix $B_\eta = \sum (\varepsilon^T \ell_i)\tilde{W}_i$ is $\varepsilon^T NN^T\tilde{w}_{ab} = \eta^T N^T\tilde{w}_{ab}$ where $\tilde{w}_{ab}$ is the ab-th second derivative vector in the $\tilde{\theta}$ coordinates i.e. $\tilde{w}_{ab}$ is the ab-th column of $\tilde{W}$. This matrix is closely related to the effective residual curvature matrix B described in Hamilton, Watts and Bates (1982): B is obtained from B_η by replacing ε with e. Second, the final term of (9) can be written as

$$\begin{aligned} NN^T(\tau^T\tilde{W}\tau) &= N(\tau^T A\tau) \\ &= \tau^T\tilde{W}^N\tau \qquad (10) \end{aligned}$$

where A is the $(n - p) \times p \times p$ intrinsic curvature array (Bates and Watts, 1980) and $\tilde{W}^N$ is obtained from $\tilde{W}$ by projecting each second derivative vector onto $C(N)$. Third, it is easily seen that the approximation given in (9) is invariant under parameter transformations, as expected. Finally, the first term $N\eta = NN^T\varepsilon$ is simply the standard linear approximation.

It follows from the above discussion that e can be expressed informatively as

$$e \simeq N\eta - UB_{\eta}\tau - \frac{1}{2}\tau^T\tilde{W}^N\tau \qquad (11)$$

In terms of the basis U, the elements of $-B_{\eta}\tau$ are the coordinates of the projection of e onto $C(V)$, while $(\eta - \frac{1}{2}\tau^T A\tau)$ contains the coordinates in the basis provided by N of the projection of e onto $C'(V)$.

Equation (11) gives our final quadratic approximation of e. In the next section, we use this approximation to investigate the moments of e.

3. MOMENTS OF e

Since τ and η are independent, it follows immediately from (11) that, to the degree of accuracy provided by the quadratic approximation,

$$\begin{aligned} Ee &= -\frac{1}{2}E(\tau^T\tilde{W}^N\tau) \\ &= -\frac{1}{2}NN^TE(\tau^T\tilde{W}\tau) \\ &= NN^Td \end{aligned} \qquad (12)$$

where d is an $n \times 1$ vector with elements $-\sigma^2 tr(\tilde{W}_i)/2 = -\sigma^2 tr[(V^TV)^{-1}W_i]/2$, $i = 1,2,\ldots,n$. The vector d is essentially the expected difference between the linear and quadratic approximations of $f(\hat{\theta})$ (see eq. 4) so that Ee is the projection of this expected difference onto $C(N)$. The expectation of e can also be expressed in terms of A:

$$Ee = -\frac{\sigma^2}{2}N\sum_i^p a_{ii} \qquad (13)$$

where a_{ii} is an $(n - p) \times 1$ vector with elements a_{jii}, $j = 1,2,\ldots,n - p$, and a_{jii} is the i-th diagonal element of the j-th face of A. The results in equations (12) and (13) agree with those of Cox and Snell (1968) for the special case of model (1).

From the discussion at the end of section 2, it is easily seen that the three addends in equation (11) are uncorrelated so that

$$\begin{aligned} Var(e) &= Var(N\eta) + Var(UB_\eta\tau) + \frac{1}{4}Var(\tau^T\tilde{W}^M\tau) \\ &= NN^T\sigma^2 + U(EB_\eta B_\eta^T)U^T\sigma^2 + \frac{1}{4}Var(\tau^T\tilde{W}^M\tau) \end{aligned} \qquad (14)$$

The matrix B_η can be written as $B_\eta = \sum \eta_i A_i$ where A_i is the i-th face of A and η_i is the i-th component of η, $i = 1,2,\ldots,n - p$. From this it follows immediately that

$$E(B_\eta B_\eta^T) = \sigma^2 K \qquad (15)$$

where $K = \Sigma A_i^2$. Next, $Var(\tau^T\tilde{W}^M\tau) = N\,Var(\tau^TA\tau)N^T$ and the ij-th element of $Var(\tau^TA\tau)$ is $Cov(\tau^TA_i\tau, \tau^TA_j\tau) = 2\sigma^4 tr(A_iA_j)$, $i,j = 1,\ldots,n - p$. Substituting this and (15) into (14) yields

$$\mathrm{Var}(e) = NN^T\sigma^2 + UKU^T\sigma^4 + \frac{1}{2} NZN^T\sigma^4 \tag{16}$$

where Z is the $(n - p) \times (n - p)$ matrix with elements $\mathrm{tr}(A_iA_j)$.

Alternatively, Z can be expressed as

$$Z = N^T(\sum_a^p \sum_b^p \tilde{w}_{ab}\tilde{w}_{ab}^T)N \tag{17}$$

From this and the forms of the first and second terms in (16), we see that $\mathrm{Var}(e)$ is positive semi-definite so that the standard linear approximation will underestimate the variances of the residuals. The amount of underestimation depends heavily on the intrinsic curvature array A, as should be clear from an inspection of (16). However, there is some doubt about the usefulness of using the elements of A as indicators of the adequacy of the linear approximation. In terms of the basis N, the columns of A contain the coordinates of the projections of the second derivative vectors $\tilde{w}_{ab}$ onto $C'(V)$. If the basis for $C'(V)$ is changed A will change. On the other hand (16) is invariant under such changes.

In linear regression, the interpretation of the standard diagnostic plot of the residuals versus the fitted values depends on the fact that the plotted quantities are uncorrelated. The interpretation of the corresponding plot in nonlinear regression may be more difficult since the residuals e and fitted values $f(\hat{\theta})$ are generally correlated. The previous development allows for a rather straightforward determination of the nature of the dependence between e and $f(\hat{\theta})$: From (11),

$$f(\hat{\theta}) = f(\theta^*) + U\tau + UB_\eta\tau + \frac{1}{2}\tau^T\tilde{W}^N\tau \tag{18}$$

Using (18), (11) and the symmetry of the error distribution, we find that

$$\mathrm{Cov}(e,f(\hat{\theta})) = -\mathrm{Var}(UB_\eta\tau) - \frac{1}{4}\mathrm{Var}(\tau^T\tilde{W}^N\tau) \tag{19}$$

so that

$$\mathrm{Var}(e) = NN^T\sigma^2 - \mathrm{Cov}(e,f(\hat{\theta})) \tag{20}$$

Thus, the covariances between the corresponding elements of e and $f(\hat{\theta})$ are negative. These covariances will be small when the linear and quadratic approximations of $\mathrm{Var}(e)$ are close.

The results of this section clearly indicate that diagnostic methods based on the standard linear approximation can potentially fail. For example, an ordinary residual may appear to be unusually large because its expectation differs substantially from zero, or because the linear approximation of $\mathrm{Var}(e)$ does not accurately reflect its variance. A plot of the ordinary residuals may exhibit systematic features because the corresponding plot of Ee exhibits such features, or because $\mathrm{Cov}(e,f(\hat{\theta}))$ is not sufficiently small. Generally, unusual characteristics of e alone are not sufficient to infer a failing of the model or data. In the next section we suggest ways to overcome the apparent shortcomings of the ordinary residuals.

4. PROJECTED RESIDUALS

A variety of useful diagnostic can be obtained by projecting e onto selected subspaces. As a class we call these projected residuals.

Recall from (11) that the difference between the linear and quadratic approximations of e depends on $UB_\eta\tau$ and $\tau^T\tilde{W}^N\tau$. These terms account for the potential problems that may be encountered in diagnostic analyses based on linear approximations. Clearly, $UB_\eta\tau$ is in $C(U)$ and $\tau^T\tilde{W}^N\tau$ is in $C(\tilde{W}^N)$, the column space spanned by $NN^T\tilde{w}_{ab}$, $a,b = 1,2,\ldots,p$, which is a subspace of $C(N) = C'(U)$. Thus, the effect of these terms can be removed by projecting e onto $C'(U,\tilde{W}^N) = C'(U,\tilde{W}) = C'(V,W)$.

Let P_{12}, P_1 and $P_{2/1}$ denote the projection operators for $C(U,\tilde{W})$, $C(U)$ and $C(\tilde{W}^N)$, respectively. Projection operators for orthogonal subspaces will be indicated by P'. Then $P_{12} = P_1 + P_{2/1}$ and

$$P_{12}'e = P_1'e - P_{2/1}e$$

$$= NN^T\varepsilon - P_{2/1}\varepsilon \ . \qquad (21)$$

The first term of (21) is the linear approximation; the second term reflects the adjustment necessary to remove the quadratic component of e. If the columns of $\tilde{W}^N$ are "small", so that the second derivatives are unimportant, we will have $P_{12}'e \simeq e$ and nothing will be lost by considering $P_{12}'e$. On the other hand, if the second derivatives are important, the adjustment provided by (21) will be important also.

The projected residuals have several useful properties in common with the residuals from linear regression. First, we clearly have $E(P_{12}'e) = 0$. Second, the projected residuals and the fitted values are uncorrelated. This property follows since $P_{12}'e$ depends only on η which is independent of τ. Finally,

$$\mathrm{Var}(P_{12}'e) = P_{12}'\sigma^2 \qquad (22)$$

and

$$E(e^TP_{12}'e) = \sigma^2\mathrm{tr}(P_{12}') \ . \qquad (23)$$

From (22) we see that the construction of Studentized projected residuals is straightforward, while (23) shows how to construct estimates of σ^2 that are free of the bias contributed by the quadratic terms.

The projected residuals overcome many of the shortcomings of the ordinary residuals and can be interpreted in much the same way as the residuals from linear regression. For example, suppose that the response function is off by a term $g(\beta)$ so that the true response function is $f(\theta) + g(\beta)$. In this case the errors become $\varepsilon + g(\beta)$ and the projected residuals are

$$P_{12}'e = P_{12}'g(\beta) + P_{12}'\varepsilon \ .$$

As in linear regression, a plot of $P_{12}'e$ against the explanatory variables associated with $g(\beta)$ may reveal the presence of the systematic component $P_{12}'g(\beta)$.

A potentie' disadvantage of the projected residuals is that there is no longer an exact correspondence between residuals and observations. There is, however, an approximate correspondence between the projected residuals and the errors in roughly the same way that there is a correspondence between the ordinary residuals and the errors in linear regression. Suppose, for example, that the first error contains an outlier of magnitude β so that $g(\beta) = \beta b_1$ where b_1 is the first standard basis vector. Then

$$P_{12}'e = \beta P_{12}'b_1 + P_{12}'\varepsilon \ .$$

As in linear regression, the first component of $P_{12}'e$ will be inflated by an amount that is usually in excess of the amount that the remaining residuals are inflated.

5. ILLUSTRATIONS

For our first illustration, we consider the class of partially nonlinear models with response functions of the form

$$f(\theta) = X\alpha + \beta g(\gamma) \tag{24}$$

where X is a known full rank $n \times (p - 2)$ matrix, $\theta^T = (\alpha^T, \beta, \gamma)$ and β and γ are scalars. This class of response functions occurs often in practice and in the statistical literature. In particular, (24) allows for transformations of explanatory variables in linear regression.

For the response function described by (24) it is easily seen that

$$V = [X, g(\gamma), \beta g^1(\gamma)] \tag{25}$$

where $g^1(\gamma)$ is the $n \times 1$ vector with elements $\partial g_i(\gamma)/\partial\gamma$, $i = 1,2,\ldots,n$. Further, there are only two nonzero second derivative vectors, $w_{\beta\gamma} = g^1(\gamma)$ and $w_{\gamma\gamma} = \beta g^2(\gamma)$ where $g^2(\gamma)$ has elements $\partial^2 g_i(\gamma)/\partial\gamma^2$. Thus,

$$C(V,W) = C(X, g(\gamma), g^1(\gamma), g^2(\gamma)) \tag{26}$$

and

$$\begin{aligned} P'_{12}e &= P'_1 e - P_{2/1}e \\ &= P'_1 Y - P_{2/1}Y \end{aligned}$$

where P'_1 is the projection operator for $C'(V)$ *and* $P_{2/1}$ is the projection operator for $C(NN^T g^2(\gamma))$. The linear approximation will work well whenever $g^2(\gamma)$ is in or lies close to $C(V)$; that is, whenever the residuals from the regression of $g^2(\gamma)$ on V are sufficiently small. Otherwise, the adjustment $P_{2/1}e$ will be important.

This condition for the adequacy of the linear approximation is also reflected by Ee given by (12) and $\mathrm{Var}(e)$ given by (16): Evaluating (12) we find

$$Ee = -\frac{1}{2}\beta\, \mathrm{Var}_\ell(\hat{\gamma}) P'_1 g^2(\gamma) \tag{27}$$

where $\mathrm{Var}_\ell(\hat{\gamma})$ is the large sample variance of $\hat{\gamma}$, i.e. the appropriate element of $(V^T V)^{-1}\sigma^2$. It can also be established that the second and third terms of (16) will be small if $P'_1 g^2(\gamma)$ is small (see equation 17).

Our second illustration, which is primarily numerical, is based on the model

$$f(x,\theta) = \theta_1 + \theta_2(x - \theta_4) + \theta_3\{(x - \theta_4)^2 + \theta_5\}^{1/2} \tag{28}$$

and data set 3 from Ratkowsky (1983, Table 6.18). The data consist of 27 observations on radioactivity counts y at equally spaced time intervals, $x = 1,2,\ldots,27$. In this example, all derivatives are estimated by substituting the maximum likelihood estimates given by Ratkowsky (1983, Table 6.19) for unknown parameters.

For this model and data set, the difference between the linear and quadratic approximations of $\mathrm{Var}(e_i)$ is small. Generally, the quadratic part of (16) accounts for only about 3% of the total variance. Since the contribution of the quadratic terms is small, the linear approximation of $\mathrm{Var}(e_i)$ with $\hat{\sigma} = s = .3095$ was used to construct the vector $S(e)$ of Studentized ordinary residuals with elements $e_i/\hat{\sigma}(\hat{P}_1')_{ii}^{1/2}$. Here and in the remainder of this discussion, a "hat" above any quantity indicates evaluation at $\hat{\theta}$.

For reference, a scatter plot of the elements of $S(e)$ versus x is given as Figure 1 and an index plot of $\mathrm{diag}(\hat{P}_1)$ versus x is given as Figure 2. The corresponding plot of the ordinary residuals is similar to that displayed in Figure 1. Notice from Figure 2 that the first two or three cases in addition to the cases, particularly the last, that fall on the plateau of the response function will have relatively large influences on the fitted model.

A plot of $\hat{E}e$ versus x is given as Figure 3. For ease of interpretation, the elements of $\hat{E}e$ have been scaled in the same way as the elements of $S(e)$. The plot of the unscaled $\hat{E}e$ versus x is similar. The residual expectations are clearly patterned, although their expected sizes indicate that the residual biases are not likely to play a dominant role in diagnostic plots. If the experimental error were increased, however, patterns such as that in Figure 3 could become extremely important.

We next turn to the projected residuals, $P_{12}'e$. The difference between the variances of the ordinary and projected residuals is indicated in Figure 4 which is a plot of $(\mathrm{diag}(\hat{P}_1')-\mathrm{diag}(\hat{P}_{12}'))$ versus x. Figure 5 gives a plot of $(S(e) - S(\hat{P}_{12}'e))$ versus x, where $S(\hat{P}_{12}'e)$ is the vector of Studentized projected residuals constructed by using (22) and (23) $(\hat{\sigma} = .3141)$. Again, a pattern is clearly evident and the largest differences occur on the plateau of the response function. The magnitudes of the differences are, of course, large enough to produce noticeable changes in various diagnostics. For example,

Figure 1. Ratkowsky Data: Scatter plot of the Studentized residuals S(e) versus x.

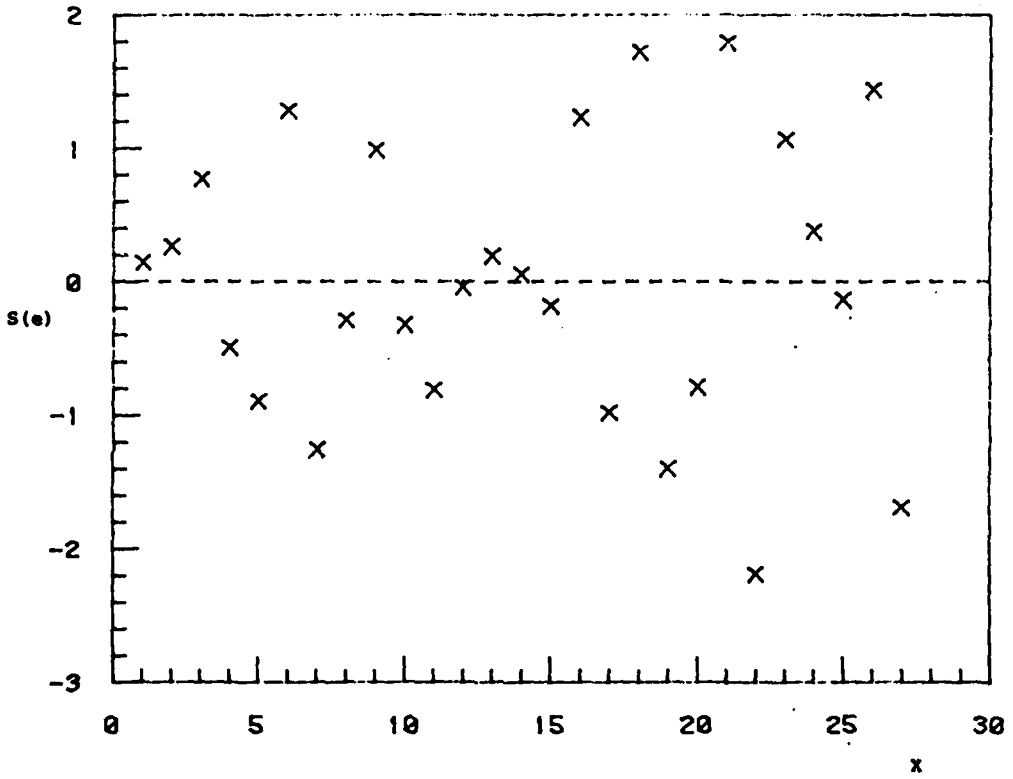


Figure 2. Ratkowsky Data: Scatter plot of $\mathrm{diag}(\hat{P}_1)$ versus x.

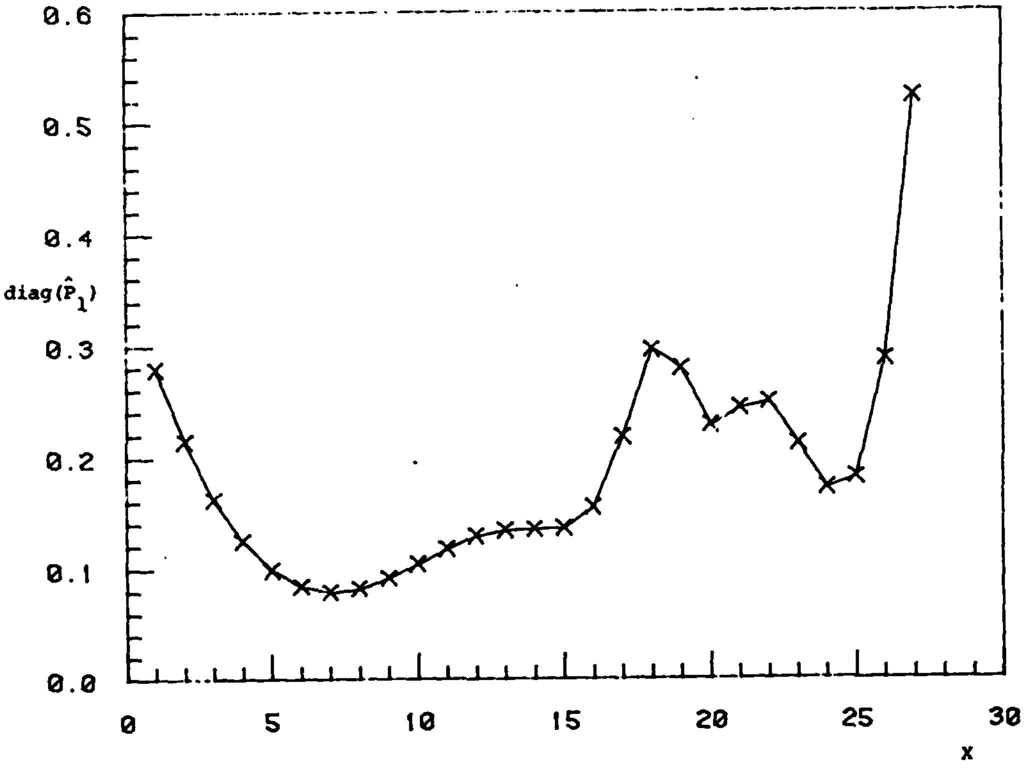


Figure 3. Ratkowsky Data: Scatter plot of the estimated residual expectation $\hat{E}e$ versus x.

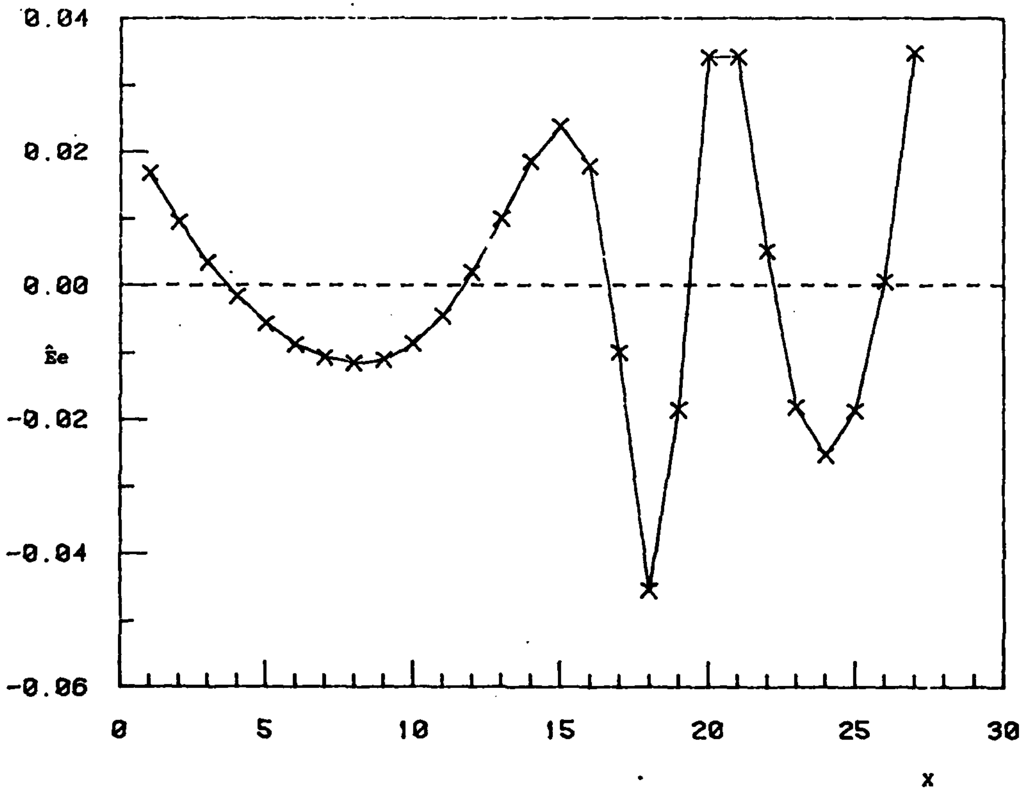


Figure 4. Ratkowsky Data: Scatter plot of
$D = (\mathrm{diag}(\hat{P}_1') - \mathrm{diag}(\hat{P}_{12}'))$ versus x.

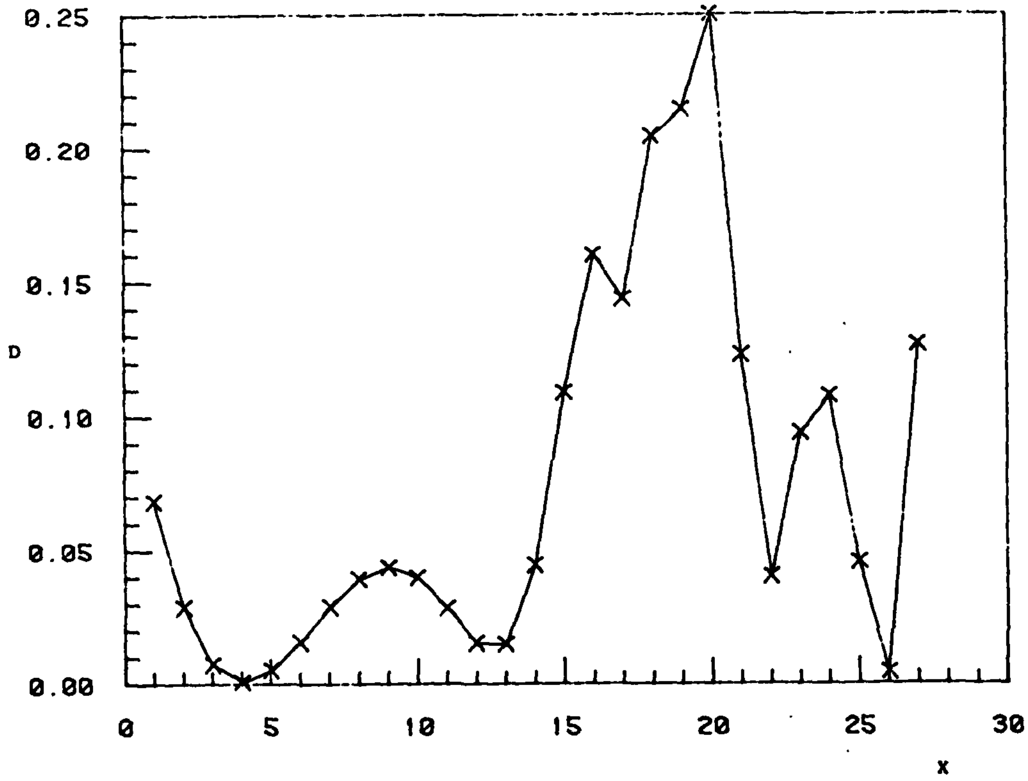
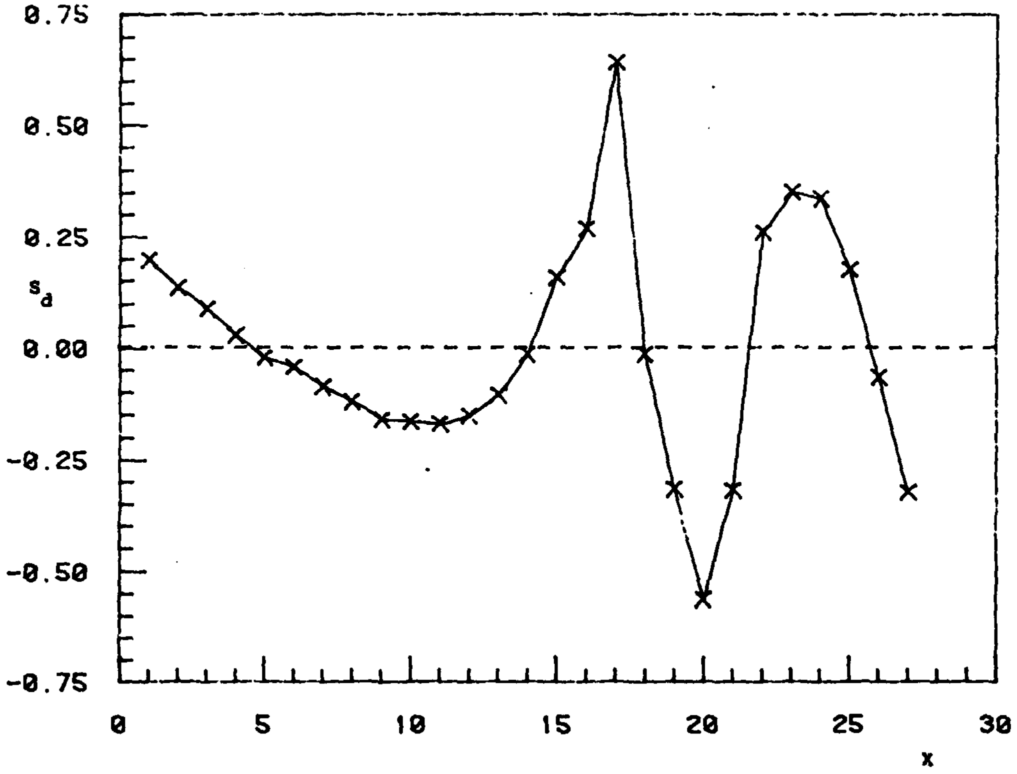


Figure 5. Ratkowsky Data: Scatter plot of $S_d = (S(e) - S(P'_{12}e))$ versus x.



the largest absolute Studentized ordinary residual is $S(e)_{22} = -2.2$ and the largest absolute Studentized projected residual is $S(\hat{P}'_{12}e)_{22} = -2.5$, which reflects an increase in the chance that case 22 may be an outlier.

The general patterns in Figures 3 and 5 are quite similar. We can explain this occurrence with a heuristic argument that also serves to point out some of the detail behind this example. First, the plot of $e - \hat{P}'_{12}e$ versus x is very similar to those in Figures 3 and 5 so that we can consider unscaled rather than Studentized residuals. Second, from (11) and (21)

$$\begin{aligned} e - P'_{12}e &= P_{12}e \\ &= P_{2/1}\varepsilon - UB_{\eta}\tau - \frac{1}{2}\tau^{T}\tilde{W}^{N}\tau \qquad (29) \end{aligned}$$

so that $P_1P_{12}e = -UB_{\eta}\tau$ which can be estimated by substituting estimates for unknown parameters. In the example at hand, the estimated elements of $UB_{\eta}\tau$ are small relative to those of $P_{12}e$. Next, of the 15 second derivative vectors, 12 are in $C(V)$. Of the remaining 3 vectors only one contributes substantially to the determination of $\hat{P}_{12}e$. Thus, $P_{12}e$ is roughly a random scalar times a single column of $\tilde{W}^{N}$. Since $Ee = E(P_{12}e)$ we can expect Figures 3 and 5 to look similar.

6. DISCUSSION

There is clearly a close relationship between the results of this paper and methods for assessing intrinsic curvature (Bates and Watts 1980). As expected, we have found that the difference between the ordinary and projected residuals is negligible when the maximum intrinsic curvature is sufficiently small. Such behavior might be taken as justification for using the maximum intrinsic curvature as a diagnostic to indicate when the difference between the ordinary and projected residuals is likely to be substantial. However, the intrinsic curvature and the projected residuals both necessitate the often tedious construction of the second derivative vectors. Once these vectors are available, the construction of the projected residuals is straightforward and can be carried out in most standard regression programs. It does not seem sensible to rely on a diagnostic that is no easier to construct than the quantity of primary interest.

In our experience, the projected residuals rarely alter the patterns of ordinary residual plots in a way that completely changes our interpretation, although we see no inherent reason why such drastic changes cannot occur with some frequency in particular applications. On the other hand, summary statistics computed from residuals often change in important ways.

Numerical problems may be encountered during the construction of the projected residuals when the second derivative vectors lie close to $C(V)$ so that the model is essentially linear. For example, when using standard regression programs to compute the projected residuals, all of the second derivative vectors will occasionally be deleted automatically because of high correlations with the columns of V.

Finally, the results described in this paper rely on the accuracy of the various quadratic approximations, of course. In principle, all results can be extended by carring the approximations to a higher order. The practical advantages of such extensions, however, are unclear.

APPENDIX

Derivation of Equation (7)

Let $L = L(\theta)$ denote the log likelihood for model (1) and without loss of generality assume that σ^2 is known. Further, let $L_r = \partial L/\partial\theta_r$, $L_{rs} = \partial L_r/\partial\theta_s$ and let $L_{rst} = \partial L_{rs}/\partial\theta_t$. The quadratic expansion of the likelihood equations $L_r(\hat{\theta}) = 0$, $r = 1,2,\ldots,p$, about the true value θ^* is

$$L_r + \sum_s \phi_s L_{rs} + \frac{1}{2}\sum_{s,t} \phi_s\phi_t L_{rst} \doteq 0 \tag{A.1}$$

where ϕ_k is the k-th component of $\phi = (\hat{\theta} - \theta^*)$.

The first term of (A.1) is simply $L_r = \sum_i^n \varepsilon_i f_i^r/\sigma^2$, $r = 1,2,\ldots,p$, or in matrix notation

$$(L_r) = V^T\varepsilon/\sigma^2 \tag{A.2}$$

For the second term, $L_{rs} = \sum_i (\varepsilon_i f_i^{rs} - f_i^r f_i^s)/\sigma^2$ so that

$$\sigma^2\left(\sum_s \phi_s L_{rs}\right) = \sum_i^n \varepsilon_i W_i\phi - V^T V\phi \tag{A.3}$$

Next,

$$L_{rst} = \sum_i^n (\varepsilon_i f_i^{rst} - f_i^{rs} f_i^t - f_i^r f_i^{st} - f_i^s f_i^{rt})/\sigma^2$$

Since the approximation is to be constructed by ignoring terms involving cubic and higher powers in the ε_i's, the first term of L_{rst} is set to zero. This gives for $r = 1,2,\ldots,p$

$$\sigma^2 \sum_{s,t}^p \phi_s\phi_t L_{rst} = -\sum_i^n \sum_{s,t}^p \phi_s\phi_t (f_i^{rs} f_i^t + f_i^r f_i^{st} + f_i^s f_i^{rt}$$

$$= -\sum_i \sum_{s,t} \phi_s\phi_t (2f_i^{rs} f_i^t + f_i^r f_i^{st})$$

or in matrix notation,

$$\sigma^2\Big(\sum_{s,t} \phi_s\phi_t L_{rst}\Big) = -\Big\{2 \sum_i (v_i^T\phi)W_i\phi + V^T(\phi^T W\phi)\Big\} \tag{A.4}$$

where v_i^T is the i-th row of V.

Finally, substituting (A.2), (A.3) and (A.4) into (A.1) and rearranging terms we find that

$$(V^TV)\phi = V^T\epsilon + \sum_i (\epsilon_i - v_i^T\phi)W_i\phi - \frac{1}{2} V^T(\phi^T W\phi)$$

or

$$V\phi = H\epsilon + V(V^TV)^{-1} \sum_i (\epsilon_i - v_i^T\phi)W_i\phi - \frac{1}{2} H(\phi^T W\phi) \tag{A.5}$$

Substituting the standard linear approximation for the ϕ's on the right side of (A.5) yields equation (7).

REFERENCES

[1] Bates, D. M. and Watts, D. G. (1980). "Relative curvature measures of nonlinearity". Journal of the Royal Statistical Scoiety B, 42, 1-25.

[2] Box, M. J. (1971). "Bias in nonlinear estimamtion". Journal of the Royai Statistical Society B, 32, 171-201.

[3] Clarke, G. P. Y. (1980). "Moments of the least squares estimators in a nonlinear regression model". Journal of the Royal Statistical Society B, 42, 227-237.

[4] Cook, R. D. and Weisberg, S. (1982). Residuals and Influence in Regression. Chapman and Hall: London.

[5] Cox, D. R. and Snell, E. J. (1968). "A general definition of residuals". Journal of the Royal Statistical Society B, 30, 248-275.

[6] Hamilton, D. C., Watts, D. G. and Bates, D. M. (1982). "Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions". Annals of Statistics, 10, 386-393.

[7] Kennedy, W. and Gentle, J. (1980). Statistical Computing. Marcel Dekker, Inc.: New York

[8] Ratkowsky, D. A. (1983). Nonlinear Regression Modeling. Marcel Dekker, Inc.: New York.

[9] Wu, C. F. (1981). "Asymptotic theory of nonlinear least squares estimation". Annals of Statistics, 9, 501-513.

RDC/CLT/ed

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS
BEFORE COMPLETING FORM

1. REPORT NUMBER 2625	2. GOVT ACCESSION NO. AD-A139 369	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) RESIDUALS IN NONLINEAR REGRESSION		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. D. Cook and C. L. Tsai		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE January 1984 13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Diagnostics, intrinsic curvature array, nonlinear regression, residuals

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We employ a quadratic expansion to investigate the behavior of the ordinary residuals in nonlinear regression. In particular, we derive quadratic approximations for the mean and variance of the ordinary residuals, and the covariances between the ordinary residuals and the fitted values. This investigation leads to the conclusion that the ordinary residuals can produce misleading results when used in diagnostic methods analogous to those for linear regression. Consequently, we suggest a new type of residual that overcomes many of the potential shortcomings of the ordinary residuals.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)